Derivation of the Onsager-Machlup function by a minimisation on the Kullback-Leibler entropy

1981 J. Phys. A: Math. Gen. 14 L385

(http://iopscience.iop.org/0305-4470/14/10/001)

J. Phys. A: Math. Gen. 14 (1981) L385–L388. Printed in Great Britain

LETTER TO THE EDITOR

Derivation of the Onsager–Machlup function by a minimisation of the Kullback–Leibler entropy

H Ito

Kakioka Magnetic Observatory, Kakioka, Yasato-machi, Niihari-gun, Ibaraki-ken, Japan

Received 5 June 1981

Abstract. The Onsager–Machlup function for general Markov processes with small parameters, inclusive of the so-called system size expansion model, is obtained by a conditional minimisation of the Kullback–Leibler entropy.

In papers by Murakami (1980) and Ito (1981a) an example was given proving that the Kullback–Leibler (KL) entropy became a useful concept for studying stochastic processes; a minimisation of the KL entropy yields the so-called equivalent or statistical linearisation of a stochastic differential equation.

The aim of the present Letter is to show that the Onsager–Machlup (OM) function is also obtained by a minimisation of the KL entropy.

As may be imagined from its definition, the OM function has an entropy-like nature; in fact Bach and Dürr (1978) have already noted that the potential part of the OM function can be interpreted as the entropy production density on the most probable path. They used properties peculiar to diffusion processes, so that it seems difficult to extend their assertion to general Markov processes. Our discussion given below is valid for both diffusion and jump Markov processes, though we have to assume the existence of a small parameter ε instead.

Let us consider a Fokker–Planck equation

$$\frac{\partial p(t,x)}{\partial t} = -\sum_{i=1}^{d} \frac{\partial[b^i(x)p(t,x)]}{\partial x^i} + \frac{\varepsilon}{2} \sum_{i,j=1}^{d} \frac{\partial^2[a_{ij}(x)p(t,x)]}{\partial x^i \, \partial x^j} \tag{1}$$

or a master equation

$$\frac{\partial p(t,x)}{\partial t} = \varepsilon^{-1} \int_{R^d} \mathrm{d}r[w(x-\varepsilon r, r)p(t, x-\varepsilon r) - w(x,r)p(t,x)] \tag{2}$$

in d-dimensional Euclidean space R^d. Here $p(t,x)$ is a transition probability density and ε is a small parameter. Model (1) is generated by a stochastic differential equation

$$\mathrm{d}X^\varepsilon(t) = b(X^\varepsilon(t))\,\mathrm{d}t + \sqrt{\varepsilon}\sigma(X^\varepsilon(t))\,\mathrm{d}w(t). \tag{3}$$

Here $w(t)$ is the d-dimensional Wiener process, and $\sigma\sigma^{\mathrm{T}} = a$. Model (2), known as the system size expansion model (van Kampen 1961, Kubo *et al* 1973), is generated by a jump-type stochastic differential equation

$$\mathrm{d}X^\varepsilon(t) = \hat{b}(X^\varepsilon(t))\,\mathrm{d}t + \varepsilon \int c(X^\varepsilon(t), u)\tilde{\nu}(\mathrm{d}t/\varepsilon\ \mathrm{d}u). \tag{4}$$

0305-4470/81/100385+04$01.50

Here $\tilde{\nu}(\cdot)=\nu(\cdot)-E[\nu(\cdot)]$, and ν is a Poisson random measure with $E[\nu(\cdot)]=\int\int \mathrm{d}t\,\mathrm{d}u/|u|^{d+1}$. The relation between (2) and (4) is as follows:

$$\int_\Gamma w(x,y)\mathrm{d}y=\int \frac{I_\Gamma[c(x,u)]\,\mathrm{d}u}{|u|^{d+1}}, \qquad \hat{b}(x)=\int \frac{c(x,u)\mathrm{d}u}{|u|^{d+1}}, \tag{5}$$

where I_Γ is an indicator function of a Borel subset Γ of R^d (Ito 1981b). We consider the stochastic differential equations (3) or (4) on a basic probability space $(\Omega, \mathscr{B}, P)$.

Let us introduce the OM function. Define a function H^ε on $R^d\times R^d$ as

$$H^\varepsilon(x,z)=\lim_{t\downarrow 0} t^{-1}\lg E_x^\varepsilon[\exp(z, X_t^\varepsilon-x)], \tag{6}$$

where E_x^ε denotes a conditional expectation with $X_0^\varepsilon=x$. For model (1)

$$H^\varepsilon(x,z)=(b(x),z)+\varepsilon(z,a(x)z)/2 \tag{7}$$

and for model (2)

$$H^\varepsilon(x,z)=\varepsilon^{-2}\int(\exp(z,u)-1)w(x,u/\varepsilon)\,\mathrm{d}u. \tag{8}$$

A function $L^\varepsilon(x,u)$ on $R^d\times R^d$ is defined as the Legendre transformation on H^ε:

$$L^\varepsilon(x,u)=\sup_{z\in R^d}\{(z,u)-H^\varepsilon(x,z)\}. \tag{9}$$

Note $H^\varepsilon(x,z)=H(x,\varepsilon z)/\varepsilon$, and $L^\varepsilon(x,u)=L(x,u)/\varepsilon$. Note also that L for model (2) is explicitly written as

$$L(x,u)=(u-b(x), a^{-1}(x)(u-b(x)))/2. \tag{10}$$

Here, and in the following, the superscript ε is suppressed when $\varepsilon=1$. We call $L(\phi_t,\dot{\phi}_t)$ the OM function for a given smooth curve ϕ because the probability of moving along ϕ is approximately $\exp(-\int_0^T L(\phi_t,\dot{\phi}_t)\mathrm{d}t/\varepsilon)$, or more precisely

$$\lim_{\delta\downarrow 0}\lim_{\varepsilon\downarrow 0}\varepsilon\lg P(|X_t^\varepsilon-\phi_t|<\delta \text{ for all } t\in[0,T])=-\int_0^T L(\phi_t,\dot{\phi}_t)\,\mathrm{d}t \tag{11}$$

(Ventsel 1976, Ventsel and Freidlin 1980, Ito 1979).

From (10) the OM function for model (1) is shown to have the form

$$L(\phi_t,\dot{\phi}_t)=(\dot{\phi}_t-b(\phi_t), a^{-1}(\phi_t)(\dot{\phi}_t-b(\phi_t)))/2. \tag{12}$$

It will be interesting to compare this with the exact OM function in Riemannian space (Takahashi and Watanabe 1980, Fujita and Kotani 1981):

$$L_{\text{exact}}(\phi_t,\dot{\phi}_t)=L^\varepsilon(\phi_t,\dot{\phi}_t)+\tfrac{1}{2}\operatorname{div}(\phi_t)-\tfrac{1}{12}\varepsilon R(\phi_t), \tag{13}$$

where div is the Riemannian divergence, and R is the scalar curvature derived from the metric tensor (a_{ij}^{-1}). Expression (12) agrees with (13) in the lowest order of ε. A similar remark holds for model (2) (Langouche *et al* 1981).

Let μ, ν be probability measures on $(\Omega,\mathscr{B})$. The KL entropy $S(\mu,\nu)$ is defined by

$$S(\mu,\nu)=\int_\Omega \mathrm{d}\mu(\omega)\lg \mathrm{d}\mu/\mathrm{d}\nu(\omega), \tag{14}$$

where $d\mu/d\nu$ denotes the Radon–Nikodym derivative, and we have assumed μ is absolutely continuous with respect to ν. It is well known that $S(\mu, \nu) \geq 0$ and $=0$ if and only if $\mu = \nu$.

Suppose that we are given a smooth curve ϕ. Take t such that $0<t<\infty$ and the probability measure ν as $\nu(X_t^\varepsilon = \phi_t)$. We consider minimising (14) by changing μ under the constraints

$$\phi_{t+h} - \phi_t = E_\mu[X_{t+h}^\varepsilon - X_t^\varepsilon], \tag{15}$$

where E_μ represents the integration with respect to μ. Such a minimum of $S(\mu, \nu)$ is denoted by $S(*, \nu)$. We allow μ to range over all measures by imposing an additional constraint

$$E_\mu[1] = 1. \tag{16}$$

Using the Lagrange multiplier method, we only have to minimise

$$\begin{aligned} \hat{S} &= S(\mu, \nu) - (\alpha, E_\mu[X_{t+h}^\varepsilon - X_t^\varepsilon] - (\phi_{t+h} - \phi_t)) - \beta(E_\mu[1] - 1) \\ &= E_\nu[g \lg g - (\alpha, X_{t+h}^\varepsilon - X_t^\varepsilon) g - \beta g] + (\alpha, \phi_{t+h} - \phi_t) + \beta \end{aligned} \tag{17}$$

without any constraints. Here $g(\omega) = d\mu/d\nu(\omega)$, and α, β are multipliers. Since $y \lg y - \gamma y$ attains its minimum when $y = \exp(\gamma - 1)$, so does $\hat{S}$ when

$$g(\omega) = \exp[\beta - 1 + (\alpha, X_{t+h}^\varepsilon - X_t^\varepsilon)]. \tag{18}$$

Substituting (18) into (15) and (16), we have

$$\phi_{t+h} - \phi_t = (\partial/\partial\alpha) \lg E_\nu[\exp(\alpha, X_{t+h}^\varepsilon - X_t^\varepsilon)], \tag{19}$$

$$\beta - 1 = -\lg E_\nu[\exp(\alpha, X_{t+h}^\varepsilon - X_t^\varepsilon)]. \tag{20}$$

Using (15) and (18)–(20), we observe that

$$S(*, \nu) = (\alpha, \phi_{t+h} - \phi_t) - \lg E_\nu[\exp(\alpha, X_{t+h}^\varepsilon - X_t^\varepsilon)]. \tag{21}$$

Take the limit $h \downarrow 0$. Assuming $\lim_{h\downarrow 0}$ and E_ν are exchangeable, we have from (6) and (21)

$$\lim_{h\downarrow 0} h^{-1} S(*, \nu) = (\alpha, \dot{\phi}_t) - H^\varepsilon(\phi_t, \alpha) \tag{22}$$

and from (6) and (19)

$$\dot{\phi}_t = (\partial/\partial\alpha) H^\varepsilon(\phi_t, \alpha). \tag{23}$$

Hence by virtue of (9)

$$\lim_{h\downarrow 0} h^{-1} S(*, \nu) = L^\varepsilon(\phi_t, \dot{\phi}_t). \tag{24}$$

The right-hand side of (24) may be regarded as minimum entropy production under the constraint that the average velocity is equal to $\dot{\phi}$:

$$\dot{\phi}_t = \lim_{h\downarrow 0} h^{-1} E_\mu[X_{t+h}^\varepsilon - X_t^\varepsilon] \tag{25}$$

(see (15)). Equation (24) claims that it is just the OM function.

References

Bach A and Dürr D 1978 *Phys. Lett.* **69A** 244–6

Fujita T and Kotani S 1981 *J. Math. Kyoto University* to appear

Ito H 1979 *Suriken Kokyuroku* **367** 134–52

—— 1981a *Prog. Theor. Phys.* **66**

—— 1981b *J. Math. Phys.* submitted

van Kampen N G 1961 *Can. J. Phys.* **39** 551–67

Kubo R, Matsuo K and Kitahara K 1973 *J. Stat. Phys.* **9** 51–96

Kullback S 1959 *Information Theory and Statistics* (New York: Wiley)

Langouche F, Roekaerts D and Tirapegui E 1981 *Preprint KUL-81/6*

Murakami T 1980 *Suriken Kokyuroku* **405** 94–111

Takahashi Y and Watanabe S 1980 *Proc. LMS Symp. on Stochastic Integrals at Durham* to appear

Ventsel A D 1976 *Theor. Prob. Appl.* **21** 227–42, 449–512

Ventsel A D and Freidlin M I 1980 *Fluctuations in Dynamical Systems under the Influence of Small Random Perturbation* (Moscow: Nauka) in Russian